

Optimization and Analysis of Probabilistic Caching in N -tier Heterogeneous Networks

Kuikui Li, Chenchen Yang, Zhiyong Chen, Meixia Tao

Abstract

In this paper, we study the probabilistic caching for an N -tier wireless heterogeneous network (HetNet) using stochastic geometry. A general and tractable expression of the successful delivery probability (SDP) is first derived. We then optimize the caching probabilities for maximizing the SDP in the high signal-to-noise ratio (SNR) region. The problem is proved to be convex and solved efficiently. We next establish an interesting connection between N -tier HetNets and single-tier networks. Unlike the single-tier network where the optimal performance only depends on the cache size, the optimal performance of N -tier HetNets depends also on the BS densities. The performance upper bound is, however, determined by an equivalent single-tier network. We further show that with even caching probabilities regardless of content popularities, to achieve a target SDP, the BS density of a tier can be reduced by increasing the cache size of the tier when the cache size is larger than a threshold; otherwise the BS density and BS cache size can be increased simultaneously. It is also found analytically that the BS density of a tier is *inverse* to the BS cache size of the same tier and is *linear* to BS cache sizes of other tiers.

Index Terms

Caching, Wireless HetNets, Density-Caching Tradeoff, Power-Caching Tradeoff

I. INTRODUCTION

The global mobile data traffic is estimated to increase to 30.6 exabytes per month by 2020, an eightfold growth over 2015, and the contribution by video is foreseen to increase from 55%

The authors are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, 200240, P. R. China. Z. Chen is also with the Cooperative Medianet Innovation Center, Shanghai, China. Email: {kuikuili, zhanchifeixiang, zhiyongchen, mxtao}@sjtu.edu.cn. Z. Chen is corresponding author.

in 2015 to 75% in 2020 [1]. To address this mobile data tsunami and hence meet the capacity requirement for the future 5G network [2], an effective and promising candidate solution is to deploy a dense network with heterogeneous base stations (BSs), such as macro BSs, relays, femto BSs and pico BSs [3]. The heterogeneous network (HetNet) can provide higher throughput and spectral efficiency. In the meantime, it also faces two challenges. One is the tremendous burden on the backhaul link due to the explosive demand for video contents during the peak time and the other is high CAPEX and OPEX due to the denser BSs.

Recently, caching popular contents at base stations has been introduced as a promising technique to offload mobile data traffic in cellular networks [4], [5]. Unlike the communication resource, the caching resources are abundant, economical, and sustainable. By taking advantage of the abundance of the caching resource in wireless networks, significant gains in network capacity can be expected [6], which enables caching to be one of essential elements of a wireless network [7].

The aim of this work is to study how caching can address the aforementioned challenges in a multi-tier HetNet. In specific, we first would like to find out what is the optimal cache placement strategy in order to alleviate the traffic burden in backhaul links to the minimum. Second, we would like to find out if the deployment cost of dense BSs can be traded by BS cache storage, and if so, what are the tradeoffs and what conditions must be met in order for it to happen.

A. Related Work

Caching has the potential to alleviate the heavy burden on the capacity-limited backhaul link and also improves user-perceived experience [8]. Utilizing the tool of stochastic geometry [9], the work [10] formulates the caching problem in a scenario where small BSs are distributed according to a homogeneous Poisson Point Process (HPPP). The authors in [11] consider a two-tier HetNet and derive a closed-form expression of the outage probability by jointly considering spectrum allocation and storage constraints. In [12], the authors consider a 3-tier HPPP-based HetNet with caching and theoretically elaborate the average ergodic rate, outage probability and delay. Considering an HPPP-based cache-enabled small cell network, a closed form expression of the outage probability and the optimal BS density to achieve a target hit probability are derived in [13]. The work [14] proposes a cluster-centric small cell network and designs cooperative transmission scheme to balance transmit diversity and content diversity. It is worth noting that

these works mainly focus on the performance analysis of cache-enabled wireless networks for given caching strategies, such as caching the most popular contents.

Caching strategy is an important issue for cache-enabled wireless networks. Previous works on the optimal caching strategy design can be classified into two trends based on whether channel fading and interference are considered. The early trend focuses on the connection topology only while ignoring channel fading and interference. The authors in [15] formulate a cache placement problem in distributed helper stations to minimize the average download delay with both uncoded and coded caching. In [16], a joint routing and caching design problem is studied to maximize the content requests served by small BSs. By reducing the NP-hard optimization problem to a variant of the facility location problem, algorithms with approximation guarantees are established. The second trend takes into account channel fading and interference for caching optimization by mostly utilizing the tool of stochastic geometry. The work [17] proposes an optimal randomized caching policy to maximize the total hit probability and overviews different coverage models to evaluate the performance. The works [18], [19] optimize the probabilistic caching strategy to maximize the successful download probability in small cell networks. Further, a closed-form expression for the optimal caching probabilities is obtained in the noise-limited scenario in [19]. In [20], a greedy algorithm is proposed to find the optimal caching strategy to minimize the average bit error rate. Further, a tradeoff between file diversity gain and channel diversity gain to reach the optimal caching placement is revealed.

Recently, caching strategy optimization is extended to wireless heterogeneous networks. The combination of the optimal caching strategy and the network heterogeneity bring more gains in network capacity. Using stochastic geometry, the work [21] investigates the optimal caching policy of helper stations to maximize the success probability and area spectral efficiency, respectively, in a two-tier HetNets. Based on [17], the work [22] considers different types of BSs with different cache capacities. The cache optimization problem for the first type of BSs is solved by assuming that the placement strategy for other types of BSs is given. The joint optimization problem for all types of BSs is yet not considered. In general, the joint optimization for cache placements in different tiers of a HetNet is very challenging due to the different tier association probability brought by content diversity as well as the complicated interference distribution by the nature of network heterogeneity.

Furthermore, a tradeoff between the BS density and total storage size is firstly presented in

[10], where the work consider a homogeneous network with the most popular caching strategy. Using the optimal caching scheme, [21] shows the tradeoff in a two-tier HetNet. Both works only present the tradeoff in simulation results. Deriving and analyzing the tradeoff theoretically has not been solved. In [23], the authors address the question that how much caching is needed to achieve the linear capacity scaling in the dense wireless network based on scaling law method.

B. Contributions

In this work, we first investigate the optimal probabilistic caching to maximize the successful delivery probability (SDP) in a general N -tier ($N \geq 2$) wireless cache-enabled HetNet. We next establish an interesting connection between N -tier HetNets and single-tier networks. We then address the tradeoffs between the BS caching capability and the BS density analytically based on the even caching strategy. The main contributions are summarized as follows:

- *Analyzing and optimizing the SDP for the N -tier HetNet:* We derive the tier association probability and the SDP by modeling the BS locations in the HetNet as N -tier independent HPPPs. The optimal probabilistic caching problem for maximizing the SDP is then formulated. We prove that this problem is concave in the high signal-to-noise ratio (SNR) region. The sufficient and necessary conditions for the optimal solution are then derived.
- *Highlighting the connection between N -tier HetNets and single-tier networks:* We further study the optimal caching problem in special cases, and find that the maximum SDP of single-tier networks only depends on the cache size while that of N -tier HetNets is also determined by the BS densities and transmit powers. Moreover, in the high SNR region, we prove that there exists a single-tier network such that the maximum SDP of the N -tier HetNet is upper bounded by that of the single-tier network. When all tiers of BSs have the same cache size, the N -tier HetNet performs the same as the single-tier network, regardless of the network heterogeneity.
- *Revealing the tradeoffs between the BS density and the BS cache size:* With even caching strategy, our analysis reveals that, to maintain a target SDP, the network parameters are related as follows: 1) increasing the BS caching capability can reduce the BS density when its cache size is larger than a threshold Q_e ; 2) the BS density λ_i is ***inversely proportional*** to the cache size Q_i in the same tier, i.e., $\lambda_i = \frac{K_1}{Q_i - Q_e}$. Here, i denotes the index of the tier and K_1 is independent of λ_i and Q_i . For the different tiers, we prove that the BS density

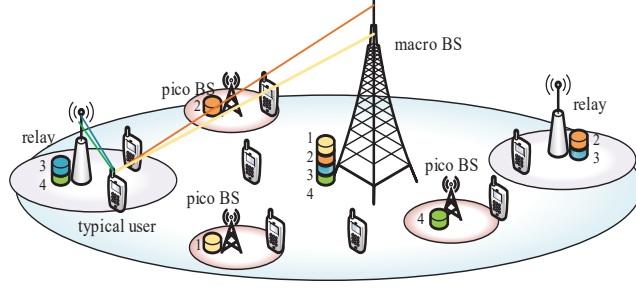


Fig. 1. An example of a 3-tier wireless cache-enabled HetNet: A typical user can obtain contents $\{3, 4\}$ from relay and contents $\{1, 2\}$ from macro BS.

λ_j is a **linear function** of the cache size Q_i , i.e., $\lambda_j = \frac{K_3 - K_4 Q_i}{K_5}$, for $i \neq j$, where K_3 , K_4 and K_5 are independent of λ_j and Q_i . Likewise, we reveal the similar tradeoffs between the BS transmit power and the BS cache size.

The rest of this paper is organized as follows. Section II presents the system model. The performance metric is analyzed in Section III. In Section IV, we formulate and solve the optimal caching problem. Then the impacts and tradeoffs of the network parameters are shown in section V. The numerical and simulation results are presented in Section VI, and the conclusions are drawn in Section VII.

II. SYSTEM MODEL

A. Network and Caching Model

We consider a general wireless cache-enabled HetNet consisting of N tiers of BSs, where the BSs in different tiers are distinguished by their transmit powers, spatial densities, biasing factors, and cache sizes. The locations of BSs in each tier are spatially distributed according to an independent HPPP, denoted as Φ_i with density λ_i for $i \in \mathcal{N} \triangleq \{1, 2, \dots, N\}$. A three-tier HetNet including macro BSs, relays and pico BSs, is illustrated in Fig. 1. Consider the downlink transmission. For the wireless channel, both large-scale fading and small-scale fading are considered. The large-scale fading is modeled by a standard distance-dependent path loss attenuation with path loss exponent β . The Rayleigh fading channel h is considered as the small-scale fading, i.e., $h \sim \mathcal{CN}(0, 1)$. Each user receiver experiences an additive noise that obeys zero-mean complex Gaussian distribution with variance σ^2 .

Consider a database consisting of M contents denoted by $\mathcal{M} \triangleq \{1, 2, \dots, M\}$, and all the contents are assumed to have equal length¹. The content popularity distribution is identical among all users, represented by $\mathbf{T} \triangleq (t_i)_{i \in \mathcal{M}}$, where each user requests the i -th content with probability t_i and $\sum_{i=1}^M t_i = 1$. The content popularity \mathbf{T} is assumed to be known a priori for cache placement. Without loss of generality, we assume $t_1 \geq t_2 \geq \dots \geq t_M$. Each BS is equipped with a cache storage. The cache capacities of N -tier BSs are denoted as $\mathcal{Q} = \{Q_1, Q_2, \dots, Q_N\}$, where each BS in the i -th tier can store at most Q_i ($Q_i < M$) contents.

When a user submits a content request, the content will be delivered directly from the local cache of a BS that has cached it. If the content is not cached in any BS, it will be downloaded from the core network through backhaul links. Since the main purpose of this paper is to optimize the caching strategy to offload the backhaul traffic, we only consider the transmission of the cached contents at BSs, same as [18].

We adopt the probabilistic caching strategy and assume all the BSs in the same tier use the same caching probabilities. Let p_{ij} denote the probability that the BSs in the i -th tier caches the j -th content. It must satisfy

$$0 \leq p_{ij} \leq 1, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \quad (1)$$

$$\sum_{j=1}^M p_{ij} \leq Q_i, \quad \forall i \in \mathcal{N}. \quad (2)$$

A realization of the probabilistic caching strategy is given in [17].

B. Probability of Tier Association

We consider a user association policy that depends on both the average received signal strength and the requested contents. Specifically, when a user requests content j , it is associated with the strongest BS among those that have cached content j from all the tiers based on the average received signal power.

We carry out our analysis for a typical user, denoted as u_0 , located at the origin without loss of generality as in [9]. Denote the distance between u_0 and the nearest BS caching content j

¹Note that the extension to the general case where contents are of different lengths is quite straightforward since the contents can be divided into chunks with equal size.

in the i -th tier by $r_{i|j}$. According to our tier association policy, the index of the tier that u_0 is associated with for content j is:

$$i(j) = \arg \max_{l \in \mathcal{N}} \{B_l S_l r_{l|j}^{-\beta}\}, \quad (3)$$

where B_l and S_l are the association bias factor and the transmit power of BSs in the l -th tier, respectively. For notation simplicity, we assume B_l is absorbed in S_l in the rest of the paper.

It is essential to determine each tier's association probability when a user requests a content. The locations of the BSs caching content j in the i -th tier can be modeled as a “thinning” HPPP Φ_{ij} with density $\lambda_i p_{ij}$. Then, according to Lemma 1 of [24], we have the following lemma.

Lemma 1. *The probability of u_0 associated with the i -th tier for content j is given by*

$$\mathcal{W}_{i|j} = \frac{\lambda_i p_{ij} S_i^{\frac{2}{\beta}}}{\sum_{l=1}^N \lambda_l p_{lj} S_l^{\frac{2}{\beta}}}. \quad (4)$$

This lemma states that the association probability $\mathcal{W}_{i|j}$ is determined directly by the density $\lambda_i p_{ij}$ and the transmit power S_i of the “thinning” HPPP Φ_{ij} .

III. PERFORMANCE ANALYSIS

In this section, we analyze the SDP for a given probabilistic caching scheme $\{p_{ij}\}_{i \in \mathcal{N}, j \in \mathcal{M}}$. We consider all the BSs operate in full loaded state and share the common bandwidth [9]. By using an orthogonal multiple access strategy within a cell, the intra-cell interference is thus not considered here and only the interference introduced by inter-tier cells and intra-tier other cells is incorporated into analysis. Given that u_0 sends a request for content j and is associated with the i -th tier. Then, the received instantaneous signal-to-interference-plus-noise ratio (SINR) of u_0 is given by

$$\text{SINR}_{i|j}(R_{i|j}) = \frac{S_i |h_{i,o}|^2 R_{i|j}^{-\beta}}{\sigma^2 + \sum_{l=1}^N \sum_{k \in \Phi_l \setminus \{n_{io}\}} S_l |h_{l,k}|^2 d_{l,k}^{-\beta}}, \quad (5)$$

where $R_{i|j}$ is the distance from u_0 to its serving BS in the i -tier tier, denoted as n_{io} , $d_{l,k}$ denotes the distance between u_0 and the k -th interfering BS in the l -th tier, $|h_{i,o}|^2 (|h_{l,k}|^2) \sim \exp(1)$ is the small-scale fading channel gain between u_0 and the serving BS (the k -th interfering BS). The delivery of content j from tier i is successful when the received SINR of u_0 is larger than a threshold τ . Thus, the SDP of content j from tier i can be calculated as

$$\mathcal{C}_{i|j} \triangleq \mathbb{E}_{R_{i|j}} [\mathbb{P} [\text{SINR}_{i|j}(R_{i|j}) > \tau]]. \quad (6)$$

By treating the locations of BSs caching content j in the i -th tier as thinning HPPPs and according to [24, Lemma 3], the probability density function (PDF) of $R_{i|j}$ is given below.

Lemma 2. *The PDF of $R_{i|j}$ is*

$$f_{R_{i|j}}(r) = \frac{2\pi p_{ij} \lambda_i}{\mathcal{W}_{i|j}} e^{-\pi \sum_{l=1}^N p_{lj} \lambda_l \left(\frac{S_l}{S_i}\right)^{\frac{2}{\beta}} r^2} r. \quad (7)$$

In the following proposition, an analytical expression for $\mathcal{C}_{i|j}$ is derived.

Proposition 1. *The SDP $\mathcal{C}_{i|j}$ is*

$$\mathcal{C}_{i|j} = \frac{2\pi p_{ij} \lambda_i}{\mathcal{W}_{i|j}} \int_0^\infty r \exp\left(-\frac{\tau r^\beta \sigma^2}{S_i}\right) \exp\left(-\pi \sum_{l=1}^N \lambda_l \left(\frac{S_l}{S_i}\right)^{\frac{2}{\beta}} (p_{lj} H(\tau, \beta) + (1-p_{lj}) D(\tau, \beta) + p_{lj}) r^2\right) dr, \quad (8)$$

where $H(\tau, \beta) \triangleq \frac{2\tau}{\beta-2} {}_2F_1(1, 1 - \frac{2}{\beta}, 2 - \frac{2}{\beta}, -\tau)$, and $D(\tau, \beta) \triangleq \frac{2}{\beta} \tau^{\frac{2}{\beta}} B(\frac{2}{\beta}, 1 - \frac{2}{\beta})$. Furthermore, ${}_2F_1(\cdot)$ denotes the Gauss hypergeometric function, and $B(x, y)$ is the Beta function defined as $\int_0^1 \mu^{x-1} (1-\mu)^{y-1} d\mu$.

Proof: Please refer to Appendix A. ■

By the total probability theorem, the average SDP for u_0 is given by

$$\mathcal{C} \triangleq \sum_{j=1}^M \sum_{i=1}^N t_j \mathcal{W}_{i|j} \mathcal{C}_{i|j}. \quad (9)$$

Substituting (8) and (4) into (9), we obtain a tractable expression of \mathcal{C} as follows

$$\mathcal{C} = \sum_{j=1}^M \sum_{i=1}^N 2\pi p_{ij} t_j \lambda_i \int_0^\infty r \exp\left(-\frac{\tau r^\beta \sigma^2}{S_i}\right) \exp\left(-\pi \sum_{l=1}^N \lambda_l \left(\frac{S_l}{S_i}\right)^{\frac{2}{\beta}} (p_{lj} H(\tau, \beta) + (1-p_{lj}) D(\tau, \beta) + p_{lj}) r^2\right) dr. \quad (10)$$

In the interference-limited scenario, where the noise power is very small compared with the interference power and hence can be neglected, the expression (10) can be simplified.

Corollary 1. *In the interference-limited scenario, i.e., $\sigma^2 \rightarrow 0$, the SDP \mathcal{C} can be simplified as*

$$\mathcal{C}' = \sum_{j=1}^M \frac{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}} p_{ij} t_j}{\sum_{l=1}^N \lambda_l S_l^{\frac{2}{\beta}} [T(\tau, \beta) p_{lj} + D(\tau, \beta)]}, \quad (11)$$

where $T(\tau, \beta) \triangleq H(\tau, \beta) - D(\tau, \beta) + 1$.

Equation (10) and (11) show a tractable expression and a closed-form expression for the SDP in the general region and high SNR region, respectively. The performance metric depends on four main factors: the number of tiers N , the caching probabilities $\{p_{ij}\}_{i \in \mathcal{N}, j \in \mathcal{M}}$, the BS densities $\{\lambda_i\}_{i \in \mathcal{N}}$ and transmit powers $\{S_i\}_{i \in \mathcal{N}}$.

IV. CACHING OPTIMIZATION AND ANALYSIS

In this section, we formulate and solve the optimal caching problem to maximize the SDP in the high SNR region. Then, we establish an interesting connection between N -tier HetNets and single-tier networks.

A. Caching Optimization

Define $\mathbf{P} \triangleq [p_{ij}]_{N \times M}$ as the caching probability matrix. The optimization problem of maximizing the SDP is formulated as

$$\begin{aligned} \text{P1 : } \quad & \max_{\mathbf{P}} \quad \mathcal{C}'(\mathbf{P}) \\ & s.t. \quad (1), (2) \end{aligned}$$

Proposition 2. *Problem P1 is a concave optimization problem.*

Proof: Please refer to Appendix B. ■

By Proposition 2, we can use the standard interior point method to solve P1 [25]. Furthermore, from (34), we can show by contradiction that the maximum SDP can be achieved only when the cache size constraint (2) holds with equality. Thus, P1 is equivalent to Problem P2 below,

$$\begin{aligned} \text{P2 : } \quad & \max_{\mathbf{P}} \quad \mathcal{C}'(\mathbf{P}) \\ & s.t. \quad 0 \leq p_{ij} \leq 1, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\ & \quad \sum_{j=1}^M p_{ij} = Q_i, \quad \forall i \in \mathcal{N} \end{aligned} \tag{12}$$

Moreover, according to (34) and (12), we have the following remark to state the impact of the cache size in each tier.

Remark 1. *For $\forall i \in \mathcal{N}$, the maximum SDP \mathcal{C}'^* increases with the cache size Q_i ($Q_i < M$).*

Let $\mathbf{P}^* = [p_{ij}^*]_{N \times M}$ denote the optimal solution of P2. By the Karush-Kuhn-Tucker (KKT) conditions, the sufficient and necessary conditions for the optimal solution of Problem P2 can be stated in the following lemma.

Lemma 3. *The optimal solution \mathbf{P}^* of Problem P2 satisfies the following sufficient and necessary conditions:*

$$p_{ij}^* = \min \left\{ \left[\frac{1}{G_i} \left(\sqrt{\frac{V_{ij}E}{\eta_i}} - \sum_{k=1, \neq i}^N G_k p_{kj}^* - E \right) \right]^+, 1 \right\} \quad (13)$$

for $\forall i \in \mathcal{N}$, $\forall j \in \mathcal{M}$, where $[x]^+ = \max\{0, x\}$, and η_i is the Lagrangian multiplier that satisfies $\sum_{j=1}^M p_{ij}^* = Q_i$ for $\forall i \in \mathcal{N}$.

Proof: Please refer to Appendix C. ■

B. Caching Optimization for Special Cases

1) *Optimization for $N=1$:* When $N = 1$, the network degrades to the single-tier network. Denote $\mathbf{p}_1 \triangleq (p_j)_{1 \times M}$ as the caching strategy and $Q < M$ as the cache size for the single-tier network. By Lemma 3, the optimal \mathbf{p}_1^* for the single-tier network is given by the following corollary².

Corollary 2. *The optimal solution to the single-tier network with $N = 1$ is*

$$p_j^* = \min \left\{ \left[\frac{1}{T(\tau, \beta)} \sqrt{\frac{t_j D(\tau, \beta)}{\eta^*}} - \frac{D(\tau, \beta)}{T(\tau, \beta)} \right]^+, 1 \right\}, \forall j \in \mathcal{M} \quad (14)$$

where η^* satisfies $\sum_{j=1}^M p_j^* = Q$ and can be found by bisection method.

Based on Corollary 1, we have the following result.

Corollary 3. *The optimal caching probability p_j^* ($j \in \mathcal{M}$) decreases with the index j , i.e., p_j^* increases with the t_j . Besides, p_j^* ($j \in \mathcal{M}$) increases with Q .*

Proof: Please refer to Appendix (D). ■

²Note that the optimal caching probability in this special case is consistent with the prior work on single-tier network in [18]. Our work extends the caching strategy optimization for a single-tier network to that for a general N -tier HetNet and contributes to presenting the impacts and essential tradeoffs of the heterogeneous network parameters.

Let $C_1'^*$ denote the maximum SDP for $N = 1$. By (11) for $N = 1$, we have the following remark.

Remark 2. *In the interference-limited region, the maximum SDP $C_1'^*$ of single-tier networks is independent of the BS density, transmit power, and only depends on the cache size. This is because the serving BS and interfering BSs have the same caching resource, and the increase in signal power is counter-balanced by the increase in interference power. Similar performance independency on the BS density and transmit power has also been observed for traditional network without cache [9].*

2) *Comparison between 1-tier and N-tier HetNets:* Define $x_j = \frac{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}} p_{ij}}{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}}}$, then we can formulate a new problem below,

$$\text{P3: } \max_{\mathbf{x}} \sum_{j=1}^M \frac{x_j t_j}{T(\tau, \beta) x_j + D(\tau, \beta)} \quad (15)$$

$$\text{s.t. } 0 \leq x_j \leq 1, \quad \forall j \in \mathcal{M} \quad (16)$$

$$\sum_{j=1}^M x_j = \frac{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}} Q_i}{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}}}. \quad (17)$$

Problem P3 is equivalent to the caching optimization problem in a single-tier network with \mathbf{x} being the caching probability vector and $Q = \frac{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}} Q_i}{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}}}$ being the equivalent cache size. We then have the following proposition which states a general relationship between the optimal performance of an N -tier HetNet and that of a single-tier network.

Proposition 3. *Let C'^* be the optimal objective of Problem P2 for the N -tier HetNet and $C_1'^*$ be that of Problem P3 for the single-tier network. Then we have*

$$C'^* \leq C_1'^*$$

with equality if $Q_i = Q, \forall i \in \mathcal{N}$.

Proof: Please refer to Appendix E. ■

Interestingly, we show that in the interference-limited region, there exists a single-tier network such that the optimal performance of an N -tier HetNet with BS cache sizes $\{Q_i\}_{i \in \mathcal{N}}$, BS densities $\{\lambda_i\}_{i \in \mathcal{N}}$, and BS transmit powers $\{S_i\}_{i \in \mathcal{N}}$ in terms of the SDP is upper bounded by that of the

single-tier network with BS cache size $Q = \frac{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}} Q_i}{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}}}$, arbitrary BS density, and arbitrary BS transmit power. And their performances are the same when $Q_i = Q, \forall i \in \mathcal{N}$.

3) *Optimization for $Q_i = Q, \forall i \in \mathcal{N}$* : According to the proof for Proposition 3 and the optimal solution for $N = 1$, we have the following corollary.

Corollary 4. *When $Q_i = Q$ for $\forall i \in \mathcal{N}$, we have $p_{ij}^* = x_j^*$ for $\forall i \in \mathcal{N}$, where x_j^* ($j \in \mathcal{M}$) follows Corollary 1.*

In the interference-limited region, the maximum SDP in this scenario is equal to that of a single-tier network with cache size Q . Based on Remark 2 and Proposition 3, we have the following remark.

Remark 3. *In the interference-limited region, when all the BSs in the N -tier HetNet have the same cache size, the maximum SDP of the HetNet is independent of the network heterogeneity in the BS density and transmit power.*

Traditionally, without caching ability at BSs, the outage probability is independent of the number of tiers, the BS densities and transmit powers in the interference-limited K -tier HetNets [24]. By introducing caching resource into the system, (11) states that the maximum SDP \mathcal{C}'^* generally depends on the number of tiers N , the BS cache size $\{Q_i\}_{i \in \mathcal{N}}$, the BS density $\{\lambda_i\}_{i \in \mathcal{N}}$ and transmit power $\{S_i\}_{i \in \mathcal{N}}$. The intuition behind this observation is that the caching resource changes the decision of a user to access a BS. In the multi-tier cache-enabled HetNet, the decision not only depends on the received SINR, but also depends on the contents cached at BSs. In order to further understand the impact of the cache size Q_i on $\mathcal{C}'(\mathbf{P})$, we theoretically illustrate the relationship between the cache size and the BS density/transmit power in the next section.

V. ANALYSIS ON NETWORK PARAMETERS UNDER EVEN CACHE

In this section, with the even caching strategy where each content is cached with equal probabilities regardless of content popularities, the equivalence between the SDP of a N -tier HetNet and that of a single-tier network is obtained. Based on this property, we further investigate the impacts of the key network parameters, i.e, the BS cache size, density and transmit power on the system performance. Finally, the tradeoffs of the BS density λ_i , transmit power S_i and cache size Q_i are found.

A. The Equivalence under Even Cache

Denote $\mathbf{p}_{1,e} = \{p_j^* = \frac{Q}{M}\}_{j \in \mathcal{M}}$ and $\mathbf{P}_e = \{p_{ij}^* = \frac{Q_i}{M}\}_{i \in \mathcal{N}, j \in \mathcal{M}}$ as the even caching strategy for the single-tier and N -tier HetNet, respectively.

Proposition 4. *In the interference-limited region, the SDP $\mathcal{C}'(\mathbf{P}_e)$ of a N -tier HetNet with the caching strategy \mathbf{P}_e equals $\mathcal{C}'_1(\mathbf{p}_{1,e})$ of the single-tier network with the cache size Q and $\mathbf{p}_{1,e}$, that is*

$$\mathcal{C}'(\mathbf{P}_e) = \mathcal{C}'_1(\mathbf{p}_{1,e}).$$

Proof: Substituting \mathbf{P}_e and $\mathbf{p}_{1,e}$ into (11) and (15), respectively, we then have

$$\mathcal{C}'(\mathbf{P}_e) = \sum_{j=1}^M \frac{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}} Q_i t_j}{\sum_{l=1}^N \lambda_l S_l^{\frac{2}{\beta}} [T(\tau, \beta) Q_l + D(\tau, \beta) M]}, \quad (18)$$

$$\mathcal{C}'_1(\mathbf{p}_{1,e}) = \sum_{j=1}^M \frac{Q t_j}{T(\tau, \beta) Q + D(\tau, \beta) M} \quad (19)$$

Let $Q = \frac{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}} Q_i}{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}}}$, $\mathcal{C}'(\mathbf{P}_e)$ is the same with $\mathcal{C}'_1(\mathbf{p}_{1,e})$. We thus have Proposition 4. \blacksquare

Actually, in the scenario that the video content popularity distribution is an uniform distribution, the even caching strategy is the optimal probabilistic caching strategy for both single-tier networks and N -tier HetNets in the interference-limited region, as illustrated in the following corollary.

Corollary 5. *When the content popularity distribution is uniform, i.e., $t_j = \frac{1}{M}$, $\forall j \in \mathcal{M}$, then the optimal caching probability of P3 is $p_j^* = \frac{Q}{M}$, $\forall j \in \mathcal{M}$, and the optimal caching strategy of P2 is $p_{ij}^* = \frac{Q_i}{M}$, $\forall j \in \mathcal{M}$, $\forall i \in \mathcal{N}$.*

Proof: When $t_j = \frac{1}{M}$, $\forall j \in \mathcal{M}$, we have $p_1^* = p_2^* = \dots = p_M^*$ based on (14). Since $\sum_{j=1}^M p_j^* = Q$, we thus have $p_j^* = \frac{Q}{M}$, $\forall j \in \mathcal{M}$. Then from Proposition 3 and Proposition 4, \mathbf{P}_e is also one optimal solution of P2. \blacksquare

B. Impacts of λ_i , S_i , Q_i , $\forall i \in \mathcal{N}$

Denote Q_e as the equivalent cache size of the N -tier HetNet, i.e., $Q_e = \frac{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}} Q_i}{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}}}$, we then have the following lemma.

Lemma 4. *The SDP $\mathcal{C}'(\mathbf{P}_e)$ increases with S_i and λ_i when $Q_i \geq \frac{\sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}} Q_j}{\sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}}}$. Otherwise, $\mathcal{C}'(\mathbf{P}_e)$ decreases with S_i or λ_i .*

Proof: From (19), we can see that $\mathcal{C}'_1(\mathbf{p}_{1,e})$ increases with Q . Let $Q = Q_e$, then we have

$$\begin{aligned} \frac{\partial Q}{\partial \lambda_i} &= \frac{S_i^{\frac{2}{\beta}} (Q_i \sum_{j=1}^N \lambda_j S_j^{\frac{2}{\beta}} - \sum_{j=1}^N \lambda_j S_j^{\frac{2}{\beta}} Q_j)}{(\sum_{j=1}^N \lambda_j S_j^{\frac{2}{\beta}})^2} = \frac{S_i^{\frac{2}{\beta}} (Q_i \sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}} - \sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}} Q_j)}{(\sum_{j=1}^N \lambda_j S_j^{\frac{2}{\beta}})^2}, \\ \frac{\partial Q}{\partial S_i} &= \frac{\frac{2}{\beta} \lambda_i S_i^{\frac{2}{\beta}-1} (Q_i \sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}} - \sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}} Q_j)}{(\sum_{j=1}^N \lambda_j S_j^{\frac{2}{\beta}})^2}. \end{aligned}$$

Obviously, when $Q_i \geq \frac{\sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}} Q_j}{\sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}}}$, we have that $\frac{\partial Q}{\partial \lambda_i} \geq 0$ and $\frac{\partial Q}{\partial S_i} \geq 0$, which also means that $\mathcal{C}'_1(\mathbf{p}_{1,e})$ increases with λ_i and S_i ³. Due to $\mathcal{C}'(\mathbf{P}_e) = \mathcal{C}'_1(\mathbf{p}_{1,e})$, we thus have Lemma 4. ■

Lemma 5. *The SDP $\mathcal{C}'(\mathbf{P}_e)$ increases with Q_i ($Q_i < M$) and the increasing speed is monotonic to $\lambda_i S_i^{\frac{2}{\beta}}$, $\forall i \in \mathcal{N}$.*

Proof: Due to $\frac{\partial Q_e}{\partial Q_i} = \frac{\lambda_i S_i^{\frac{2}{\beta}}}{(\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}})^2} > 0$, Q_e increases with Q_i , $\forall i \in \mathcal{N}$. Since $\mathcal{C}'_1(\mathbf{p}_{1,e})$ increases with Q and $\mathcal{C}'(\mathbf{P}_e)$ is equal to $\mathcal{C}'_1(\mathbf{p}_{1,e})$ with $Q = Q_e$, we thus have this lemma. ■

Remark 4. *From Lemma 4 and Lemma 5, we can observe two important features in the N -tier cache-enabled HetNet:*

- *Increasing the density or transmit power of the BSs from the tier with small cache size decreases the system performance. This somewhat surprising result is actually intuitive since such BSs (e.g., pico or femto) with small cache only provide little service but bring strong interferences to other BSs (e.g., macro or relay).*
- *If the transmit power and density of the i -th tier BSs are both the largest among all the tiers, it is the most effective to increase the performance of the network by increasing the cache size of the i -th tier BSs.*

³This is not contradict with Remark 2 in Section IV-B1, because the BS densities $\{\lambda_i\}_{i \in \mathcal{N}}$, transmit powers $\{S_i\}_{i \in \mathcal{N}}$ and cache sizes $\{Q_i\}_{i \in \mathcal{N}}$ affect the cache size Q in here based on Q_e .

C. Tradeoffs of Q_i , λ_i , and S_i , $\forall i \in \mathcal{N}$

In the preceding subsection, we described the impacts of Q_i , λ_i , and S_i on $\mathcal{C}'(\mathbf{P}_e)$. Now we will present the tradeoffs of these network parameters at a target SDP.

1) *Tradeoffs of one tier parameters:* First, we focus on the tradeoffs of parameters in the same tier. Given the target SDP $\mathcal{C}'(\mathbf{P}_e)$, the communication resource (λ_j and S_j) and the caching resource (Q_j) of all the tiers without the i -th tier, i.e., $\forall j \in \mathcal{N}$ and $j \neq i$, we can obtain Q_e based on (19), thereby gaining the tradeoffs of the i -th tier's cache size Q_i , BS density λ_i and transmit power S_i , as illustrated in the following theorem.

Theorem 1. *With the even caching strategy, given the target $\mathcal{C}'(\mathbf{P}_e)$ determined by $\mathcal{C}'_1(\mathbf{p}_{1,e})$ and the fixed values λ_j , S_j , and Q_j , $\forall j \in \mathcal{N}$ and $j \neq i$, the network parameters Q_i , λ_i , S_i satisfy the following tradeoffs*

$$\lambda_i = \frac{K_1}{Q_i - Q_e}, \text{ for given } S_i, \forall i \in \mathcal{N}, \quad (20)$$

$$S_i = \left(\frac{K_2}{Q_i - Q_e} \right)^{\frac{\beta}{2}}, \text{ for given } \lambda_i, \forall i \in \mathcal{N}, \quad (21)$$

where

$$K_1 = \sum_{j=1, \neq i}^N \lambda_j \left(\frac{S_j}{S_i} \right)^{\frac{2}{\beta}} (Q_e - Q_j), \quad (22)$$

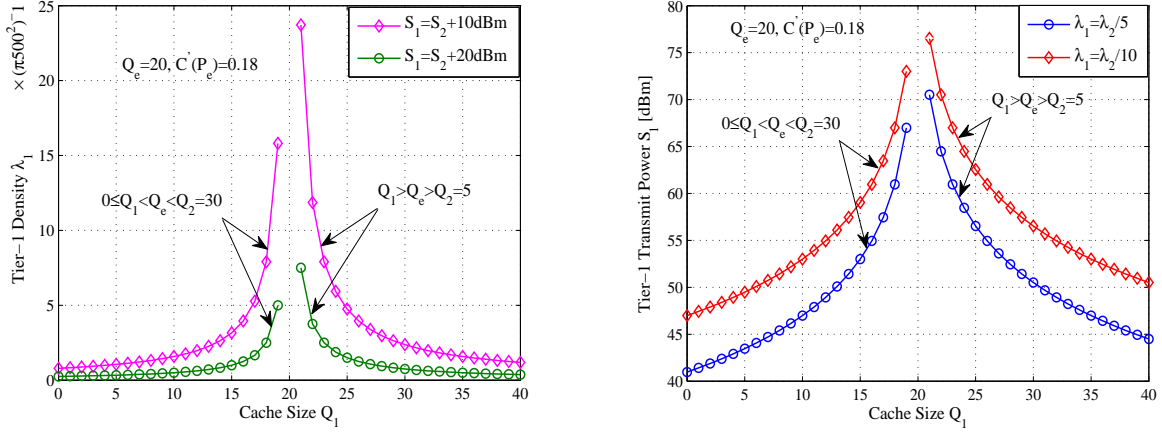
$$K_2 = \sum_{j=1, \neq i}^N \left(\frac{\lambda_j}{\lambda_i} \right) S_j^{\frac{2}{\beta}} (Q_e - Q_j). \quad (23)$$

Proof: Please refer to Appendix F. ■

Interestingly, we can observe from Theorem 1 that λ_i is inversely proportional to Q_i , while S_i is a power function of Q_i with a negative exponent $(-\beta/2)$. Accordingly, it is nature to ask that if the BS's density can be reduced by increasing the BS's caching capability. If yes, what is the condition? Thus, we have the following corollary to answer this question.

Corollary 6. *To achieve the same SDP, we have the following results for the i -th tier*

- The BS density λ_i and transmit power S_i decrease with increasing the cache size Q_i , i.e., $K_1 > 0$ and $K_2 > 0$, when $Q_i \in (Q_e, \infty)$.
- The BS density λ_i and transmit power S_i increase with increasing the cache size Q_i , i.e., $K_1 < 0$ and $K_2 < 0$, when $Q_i \in [0, Q_e)$.



(a) The tradeoffs of λ_1 and Q_1 . Two cases: (i). $\{S_1, S_2\}=\{43$ dBm, 33 dBm} (ii). $\{S_1, S_2\}=\{53$ dBm, 33 dBm}. We use $\lambda_2=\frac{5}{\pi 500^2}$ for both cases. (b) The tradeoffs of S_1 and Q_1 . Two cases: (i). $\{\lambda_1, \lambda_2\}=\{\frac{2}{\pi 500^2}, \frac{10}{\pi 500^2}\}$ (ii). $\{\lambda_1, \lambda_2\}=\{\frac{1}{\pi 500^2}, \frac{10}{\pi 500^2}\}$. We fix $S_2=33$ dBm for both cases.

Fig. 2. Under fixed $\mathcal{C}'(\mathbf{P}_e)$, the tradeoffs between the density λ_1 , transmit power S_1 and the cache size Q_1 in two-tier HetNets: $T = -10$ dB, $t_j = \frac{1/j^\gamma}{\sum_{k=1}^M 1/k^\gamma}$, $\gamma = 0.8$, $\beta = 4$, $\mathcal{C}'(\mathbf{P}_e) = 0.18$ ($Q_e=20$). According to Corollary 6, we have $K_1 < 0$, $K_2 < 0$ for $Q_1 \in [0, Q_e)$, and $K_1 > 0$, $K_2 > 0$ for $Q_1 \in (Q_e, +\infty)$.

Proof: Please refer to Appendix G. ■

It is worth mentioning that *increasing the caching capability don't always reduce the density or transmit power to achieve the same performance*. Taking $N = 2$ for example, the relationship between λ_i and Q_i is shown in Fig. 2(a) for $i = 1$. The target SDP determined by $\mathcal{C}'_1(\mathbf{p}_{1,e})$ is 0.18. Note that if a tier is deployed with larger cache size ($Q_1 > Q_e = 20$), the more the BS's cache size (Q_1) has, the less the BS density λ_i becomes, in order to maintain the same $\mathcal{C}'(\mathbf{P}_e)$. However, if the tier is deployed with small cache size ($Q_1 < Q_e = 20$), both increasing the cache size and node density can keep the same performance, as show in Fig. 2(a). This is because the increase in the tier's density of small cache size can improve its own association probabilities, but also causes the strong interference for the tier with larger cache sizes. Therefore, more users are associated with the tier of small cache size and the SDP decreases because the tier can only serves less content requests. The relationship between S_i and Q_i shown in Fig. 2(b) are similar to that of λ_i and Q_i .

2) *Tradeoffs of different tier parameters:* Similarly, we can find the tradeoffs of the parameters of different tiers, as specified in Theorem 2.

Theorem 2. When Q_i , λ_j , and S_j ($i, j \in \mathcal{N}$, $i \neq j$) satisfy the following tradeoffs for any two

different tiers, $\mathcal{C}'(\mathbf{P}_e)$ remains constant.

$$\lambda_j = \frac{K_3 - K_4 Q_i}{K_5}, \quad (24)$$

$$S_j = \left(\frac{K_3 - K_4 Q_i}{K_6} \right)^{\frac{\beta}{2}}, \quad (25)$$

for $\forall i, j \in \mathcal{N}$, $i \neq j$, where

$$K_3 = Q_e \sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}} - \sum_{k=1, \neq i, j}^N \lambda_k S_k^{\frac{2}{\beta}} Q_k, \quad (26)$$

$$K_4 = \lambda_i S_i^{\frac{2}{\beta}}, \quad (27)$$

$$K_5 = S_j^{\frac{2}{\beta}} (Q_j - Q_e), \quad (28)$$

$$K_6 = \lambda_j (Q_j - Q_e). \quad (29)$$

Proof: When $\mathcal{C}'(\mathbf{p}_{1,e})$ keeps constant, Q will be a fixed value. We consider the tradeoffs between the j -th tier parameters λ_j , S_j and the i -th tier parameter Q_i assuming that other parameters are constants. Then according to $Q_e = \frac{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}} Q_i}{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}}} = Q$, we have

$$Q_e \sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}} - \sum_{k=1, \neq i, j}^N \lambda_k S_k^{\frac{2}{\beta}} Q_k = \lambda_i S_i^{\frac{2}{\beta}} Q_i + \lambda_j S_j^{\frac{2}{\beta}} (Q_j - Q_e), \quad (30)$$

Substituting (26), (27), (28), (29) into (30), we thus have (24), (25). \blacksquare

Different with Theorem 1, **the node density λ_j is a linear function of the cache size Q_i and the transmit power S_j is a power function of Q_i with a positive exponent $\frac{\beta}{2}$** for different tiers. And the function changing law follows the following corollary.

Corollary 7. *For different tiers, we have*

- When $Q_j > Q_e$, we have $K_5 > 0$, $K_6 > 0$ and $K_3 > 0$, i.e., λ_j and S_j decrease with Q_i where $Q_i \in [0, \frac{K_3}{K_4}]$;
- When $Q_j < Q_e$ and $Q_e > \frac{\sum_{k=1, \neq i, j}^N \lambda_k S_k^{\frac{2}{\beta}} Q_k}{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}}}$, we have $K_5 < 0$, $K_6 < 0$ and $K_3 > 0$, i.e., λ_j and S_j increase with Q_i where $Q_i \in [\frac{K_3}{K_4}, +\infty)$;
- When $Q_e < \frac{\sum_{k=1, \neq i, j}^N \lambda_k S_k^{\frac{2}{\beta}} Q_k}{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}}}$, we have $K_5 < 0$, $K_6 < 0$ and $K_3 < 0$, i.e., λ_j and S_j increase with Q_i where $Q_i \in [0, +\infty)$.

Proof: According to (26)-(28), we consider the following cases:

- 1) When $K_3 > 0$, from (26), we then have $Q_e > \frac{\sum_{k=1, \neq i, j}^N \lambda_k S_k^{\frac{2}{\beta}} Q_k}{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}}}$.
- 2) When $K_5 > 0$ ($K_6 > 0$), i.e., $Q_j > Q_e$, then based on $Q_e = \frac{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}} Q_k + \lambda_j S_j^{\frac{2}{\beta}} Q_j}{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}} + \lambda_j S_j^{\frac{2}{\beta}}}$, we have $Q_e > \frac{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}} Q_k + \lambda_j S_j^{\frac{2}{\beta}} Q_e}{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}} + \lambda_j S_j^{\frac{2}{\beta}}}$, i.e., $Q_e > \frac{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}} Q_k}{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}}} > \frac{\sum_{k=1, \neq i, j}^N \lambda_k S_k^{\frac{2}{\beta}} Q_k}{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}}}$. And thus $K_3 > 0$. Due to $\lambda_j \geq 0$ and $S_j \geq 0$, we thus have $Q_i \in [0, \frac{K_3}{K_4}]$.
- 3) When $K_5 < 0$ ($K_6 < 0$) and $K_3 > 0$, i.e., $Q_j < Q_e$ and $Q_e > \frac{\sum_{k=1, \neq i, j}^N \lambda_k S_k^{\frac{2}{\beta}} Q_k}{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}}}$, we thus have $Q_i \in [\frac{K_3}{K_4}, +\infty)$ due to $\lambda_j \geq 0$ and $S_j \geq 0$.
- 4) When $K_3 < 0$, i.e., $Q_e < \frac{\sum_{k=1, \neq i, j}^N \lambda_k S_k^{\frac{2}{\beta}} Q_k}{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}}}$, we then have $Q_e < \frac{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}} Q_k}{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}}}$. Further, according to $Q_e = \frac{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}} Q_k + \lambda_j S_j^{\frac{2}{\beta}} Q_j}{\sum_{k=1, \neq j}^N \lambda_k S_k^{\frac{2}{\beta}} + \lambda_j S_j^{\frac{2}{\beta}}}$, we have $Q_j < Q_e$, i.e., $K_5 < 0$ ($K_6 < 0$). Thus, $Q_i \in [0, +\infty)$.

We thus have this corollary. ■

Remark 5. To keep the target SDP, the tradeoffs of the parameters of two different tier are summarized as two cases:

- For the tier with cache size $Q_i > Q_e$, the node density and transmit power can be reduced when another tier's cache size increases;
- For the tier with cache size $Q_i < Q_e$, the node density and transmit power increases when another tier's cache size increases.

VI. SIMULATION AND NUMERICAL RESULTS

In this section, simulation and numerical results are provided to evaluate the performance of the proposed caching strategy, and present the impacts and tradeoffs of the network parameters. We consider a two-tier cache-enabled HetNet where the BS densities $\{\lambda_1, \lambda_2\} = \{\frac{1}{\pi 500^2}, \frac{5}{\pi 500^2}\}$, the transmit powers $\{S_1, S_2\} = \{53, 33\}$ dBm and the cache sizes $\{Q_1, Q_2\} = \{40, 10\}$. We consider the interference-limited scenario and set the number of contents $M = 200$, the path loss exponent $\beta = 4$, and the SINR threshold $\tau = -10$ dB. In the simulation, the BSs are deployed according to independent HPPPs in an area of $5000 m \times 5000 m$, the content popularity distribution is modeled as the Zipf distribution, i.e., the popularity of the j -th ranked content is given by $t_j = \frac{1/j^\gamma}{\sum_{k=1}^M 1/k^\gamma}$, where $\gamma \geq 0$ characterizes the skewness of the popularity distribution, and the

optimal caching strategy is obtained by using interior point method. Here, we use $\gamma = 0.8$. These parameters will not change unless specified otherwise.

A. Optimal Caching Strategy

Comparison with other caching strategies. As shown in Fig. 3, the optimal probabilistic caching always outperforms the two considered baselines. The first baseline is popular caching strategy where each BS only caches the most popular contents, and the second is even caching strategy where each BS in all tiers caches each content randomly with even probabilities, i.e., $p_{ij} = \frac{Q_i}{M}$, $\forall i \in \mathcal{N}$, $\forall j \in \mathcal{M}$. Caching the most popular contents can only perform as good as the optimal probabilistic caching with highly skewed content popularity, e.g., $\gamma > 1$. The reason for this is that the majority of user requests only focuses on those few popular contents. The performance of even caching strategy is not affected by γ since it caches each content with even probabilities. The theoretical results for the proposed optimal caching scheme are also validated by simulation in Fig. 3.

Optimal caching probabilities for different network parameters. In Fig. 4, we consider four cases:

- 1) Case 1: $\{Q_1, Q_2\} = \{30, 5\}$, $\{\lambda_1, \lambda_2\} = \{\frac{1}{\pi 500^2}, \frac{10}{\pi 500^2}\}$, $\{S_1, S_2\} = \{43, 33\}$ dBm;
- 2) Case 2: $\{Q_1, Q_2\} = \{40, 10\}$, $\{\lambda_1, \lambda_2\} = \{\frac{1}{\pi 500^2}, \frac{10}{\pi 500^2}\}$, $\{S_1, S_2\} = \{43, 33\}$ dBm;
- 3) Case 3: $\{Q_1, Q_2\} = \{30, 5\}$, $\{\lambda_1, \lambda_2\} = \{\frac{1}{\pi 500^2}, \frac{5}{\pi 500^2}\}$, $\{S_1, S_2\} = \{43, 33\}$ dBm;
- 4) Case 4: $\{Q_1, Q_2\} = \{30, 5\}$, $\{\lambda_1, \lambda_2\} = \{\frac{1}{\pi 500^2}, \frac{5}{\pi 500^2}\}$, $\{S_1, S_2\} = \{53, 40\}$ dBm.

The content popularity based on Zipf's law is also illustrated in Fig. 4. We can observe that the optimal caching probabilities highly depend on the network parameters, e.g., cache size, content popularity, transmit power and BS density. Caching the most popular contents is not always the best solution. Specifically, the optimal caching probability is positively related to the content popularity. The results also show that adding cache size has a more significant impact on the optimal caching probabilities than increasing the BS density and transmit power.

B. Impacts of network parameters

Impacts of cache size. Fig. 5(a) illustrates that increasing the cache size can greatly improve the maximum SDP, which verifies the validity of Remark 1. And increasing λ_1 or S_1 can further

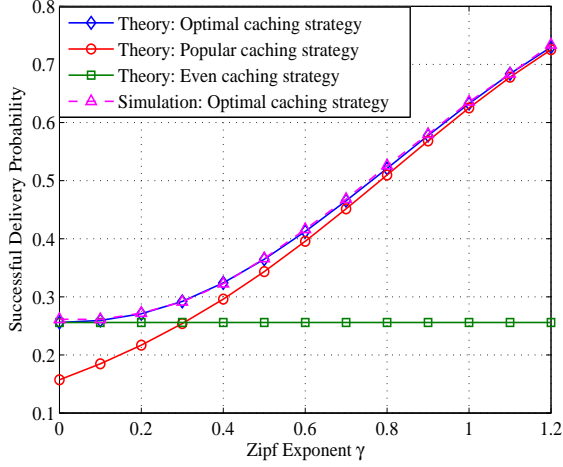


Fig. 3. SDP with different caching strategies.

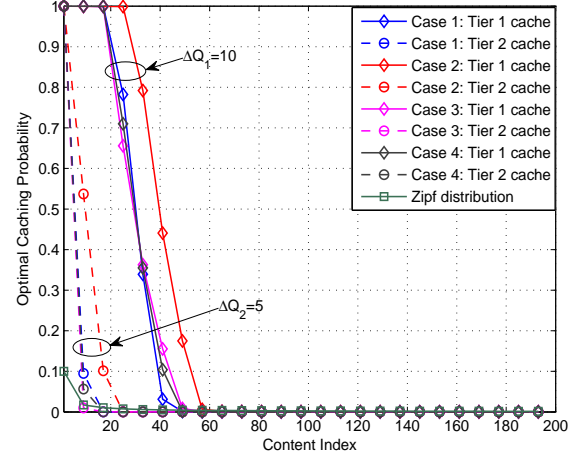


Fig. 4. The optimal caching probability p_{ij}^* for different storage sizes and transmit powers.

improve the optimal performance, since tier-1 BSs have higher association ability and more cache resource and more user requests can be served.

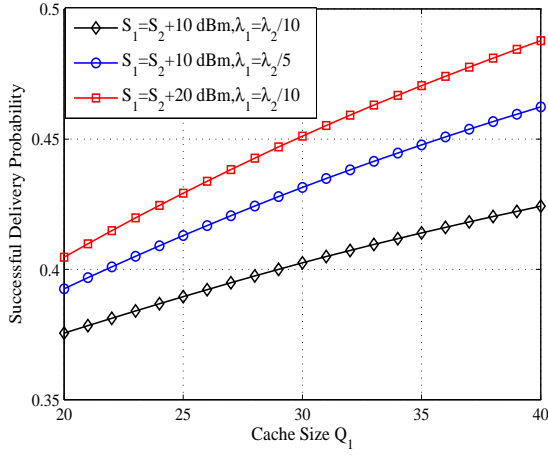
Impacts of BS densities and transmit powers. It can be observed from Fig. 5(b) that the optimal SDP decreases with the density λ_2 (with $Q_2 = 10$) and increases with λ_1 (with $Q_1 = 40$). This means deploying more BSs with smaller cache size decreases the optimal performance. This is expected because BSs with smaller cache size (e.g., pico or femto) only provide less service to satisfy user requests and they bring strong interference to other BSs (e.g. macro or relay stations). In addition, the impacts of BS transmit powers are similar to that of BS densities, which is illustrated in Fig. 5(c) and Fig. 5(d).

C. The special cases

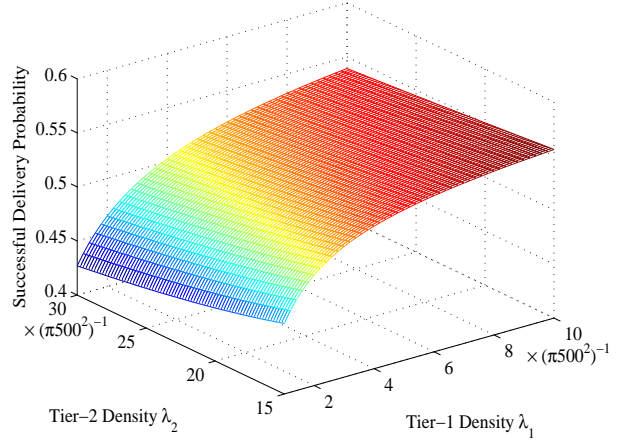
Comparison between 1-tier networks and N -tier HetNets. In Fig. 6, we consider four scenarios:

- 1) 1-tier network with $Q = 20$;
- 2) 2-tier HetNets with $\{Q_1, Q_2\} = \{20, 20\}$;
- 3) 2-tier HetNets with $\{Q_1, Q_2\} = \{25, 10\}$;
- 4) 2-tier HetNets with $\{Q_1, Q_2\} = \{28, 4\}$.

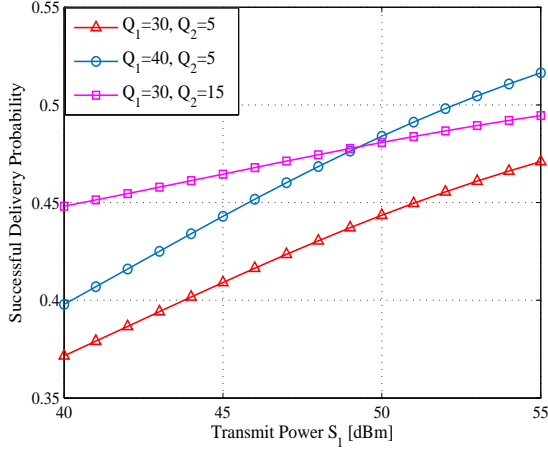
For scenarios 2 – 4), the cache sizes satisfy $\frac{\sum_{i=1}^2 \lambda_i S_i^{\frac{2}{\beta}} Q_i}{\sum_{i=1}^2 \lambda_i S_i^{\frac{2}{\beta}}} = Q = 20$. It is seen that the maximum successful delivery probabilities of 2-tier HetNets in scenarios 2 – 4) are always upper bounded by that of the 1-tier network in scenario 1). In addition, the 2-tier HetNet of scenario 2) achieves



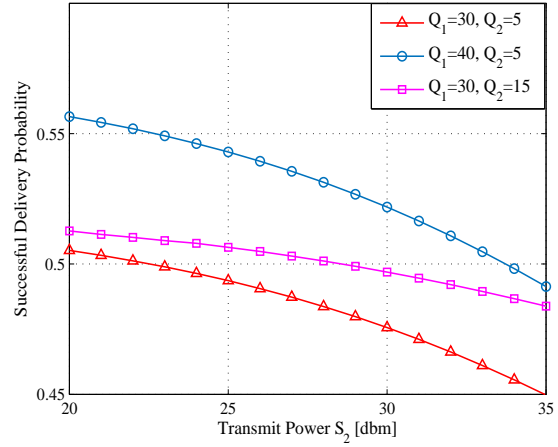
(a) Cache size Q_1 . Here, we use $\{\lambda_1, \lambda_2\} = \{\frac{1}{\pi 500^2}, \frac{5}{\pi 500^2}\}$ or $\{\frac{1}{\pi 500^2}, \frac{10}{\pi 500^2}\}$, $\{S_1, S_2\} = \{43 \text{ dBm}, 33 \text{ dBm}\}$ or $\{53 \text{ dBm}, 33 \text{ dBm}\}$.



(b) Densities $\{\lambda_1, \lambda_2\}$.



(c) Tier-1 transmit power S_1 .



(d) Tier-2 transmit power S_2 .

Fig. 5. The impacts of cache size Q_1 , densities $\{\lambda_1, \lambda_2\}$, transmit powers $\{S_1, S_2\}$ on the maximum SDP \mathcal{C}^* .

the same performance as the 1-tier network of scenario 1). These results verify our conclusions in Proposition 3 and Remark 3.

Impacts of BS density for special cases. We consider the following four special scenarios in Fig. 7.

- 1) 1-tier network with $Q = 20, S = 33 \text{ dBm}$;
- 2) 2-tier HetNets with $Q_1 = Q_2 = 15, \{S_1, S_2\} = \{53, 33\} \text{ dBm}, \lambda_2 = \frac{5}{\pi 500^2}$;
- 3) 2-tier HetNets with $\{Q_1, Q_2\} = \{15, 5\}, \{S_1, S_2\} = \{73, 33\} \text{ dBm} (S_1 \gg S_2), \lambda_2 = \frac{5}{\pi 500^2}$;
- 4) 2-tier HetNets with $\{Q_1, Q_2\} = \{15, 5\}, S_1 = S_2 = 33 \text{ dBm}, \lambda_2 = \lambda_1$.

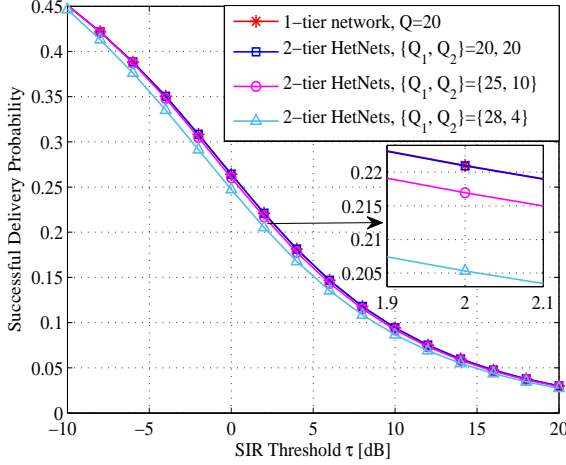
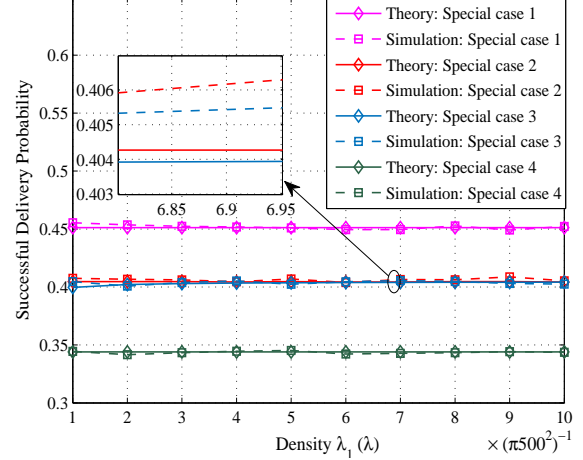


Fig. 6. Comparison between 1-tier networks and 2-tier HetNets.

Fig. 7. The impacts of BS density on the maximum SDP C'^* for four special cases.

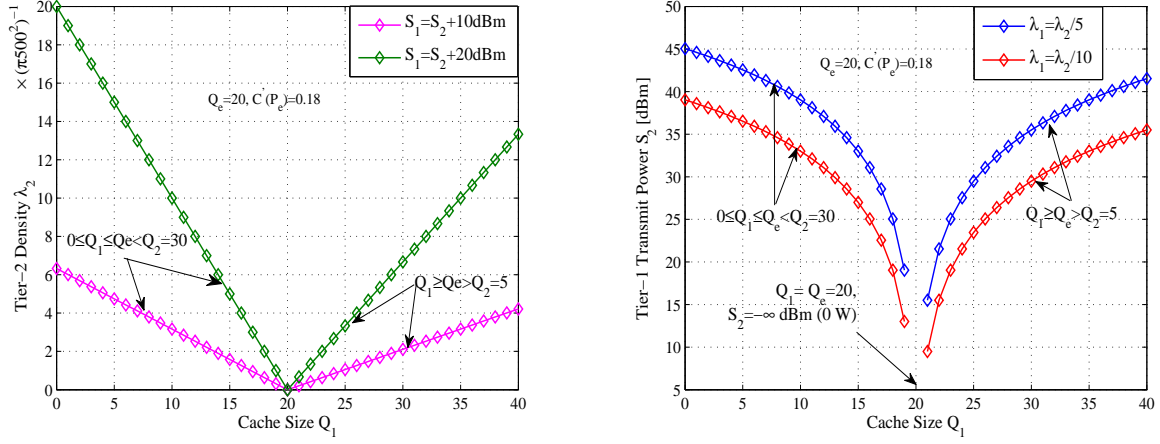
We can see that the optimal performance is almost independent of the BS density in the considered four special cases, which also verifies our conclusions for scenario 1) and 2) in Section IV-B. The optimal performance of scenario 2) is almost the same as that of scenario 3) because both them are almost entirely determined by the tier with cache size $Q = 15$.

D. Tradeoffs of network parameters

Similar to Fig. 2, we present the tradeoffs for different tiers here. As shown in Fig. 8, to keep fixed $C'(\mathbf{P}_e)$, increasing the cache size Q_1 of the first tier can't always bring the reduction of the density λ_2 or transmit power S_2 of the second tier, and the crucial condition is that the cache size of the second tier satisfies $Q_2 > Q_e$. From Fig. 8(a), we can see that the slope that λ_2 varies with Q_1 is positively related to S_2 . Similarly, the slope that S_2 varies with Q_1 is also positively related to λ_1 , as illustrated in Fig. 8(b).

VII. CONCLUSION

In this paper, we have contributed to developing a general N -tier wireless cache-enabled HetNet framework and have proposed an optimal caching probabilities scheme for maximizing the SDP. Our analysis results show that the optimal performance of such network not only depends on the cache size, but also depends also on the BS densities. Numerical results show that the proposed optimal caching strategy achieves a significant performance gain.



(a) The tradeoffs of λ_2 and Q_1 . Two cases: (i). $\{S_1, S_2\}=\{43, 33\}$ dBm (ii). $\{S_1, S_2\}=\{53, 33\}$ dBm. We use $\lambda_1=\frac{1}{\pi 500^2}$ for both cases. (b) The tradeoffs of S_2 and Q_1 . Two cases: (i). $\{\lambda_1, \lambda_2\}=\{\frac{2}{\pi 500^2}, \frac{10}{\pi 500^2}\}$ (ii). $\{\lambda_1, \lambda_2\}=\{\frac{1}{\pi 500^2}, \frac{10}{\pi 500^2}\}$. We set $S_1=53$ dBm for both cases.

Fig. 8. Under fixed $C'(\mathbf{P}_e)$, the tradeoffs between the density λ_2 , the transmit power S_2 and the cache size Q_1 in two-tier HetNets. $C'(\mathbf{P}_e) = 0.18$ ($Q_e=20$). According to Corollary 7, we have $K_3 > 0$, $K_5 > 0$ and $Q_1 \in [0, Q_e]$ for $Q_2 = 30 > Q_e$ and $K_3 > 0$, $K_5 < 0$, $Q_1 \in (Q_e, +\infty)$ for $Q_2 = 5 < Q_e$.

Toward analyzing the optimal performance of N-tier HetNets, we have discussed an interesting connection between N-tier HetNets and single-tier networks. Accordingly, we have then quantified the tradeoff between the BS cache size and the BS density for the even caching probabilities strategy, and show that: *i*) increasing the BS caching capability don't always reduce the BS density; *ii*) the BS density decreases with increasing the BS caching capability when the BS caching size is larger than a threshold; *iii*) the BS density is inversely proportional to the BS caching size for the same tier, while it becomes a linear function for different tiers.

APPENDIX

A. Proof of Proposition 1

According to (6) and let $I_l = \sum_{k \in \Phi_l \setminus \{n_{io}\}} S_l |h_{l,k}|^2 d_{l,k}^{-\beta}$, we then have

$$\begin{aligned}
\mathcal{C}_{i|j} &= \int_0^\infty \mathbb{P} [\text{SINR}_{i|j}(r) > \tau | r] f_{R_{i|j}}(r) dr \\
&= \int_0^\infty \mathbb{P} \left[|h_{i,o}|^2 > \tau S_i^{-1} r^\beta (\sigma^2 + \sum_{l=1}^N I_l) | r \right] f_{R_{i|j}}(r) dr \\
&\stackrel{(a)}{=} \int_0^\infty \mathbb{E}_{I_l} \left[e^{-\tau S_i^{-1} r^\beta (\sigma^2 + \sum_{l=1}^N I_l)} | r, I_l \right] f_{R_{i|j}}(r) dr \\
&= \int_0^\infty \exp(-\tau S_i^{-1} r^\beta \sigma^2) \prod_{l=1}^N \mathcal{L}_{I_l}(\tau S_i^{-1} r^\beta) f_{R_{i|j}}(r) dr, \tag{31}
\end{aligned}$$

where (a) follows $|h_{i,o}|^2 \sim \exp(1)$. For u_0 , there exist two types of interferences from the l -th tier: 1) the interference I_{lj+} from the nodes caching content j in the l -th tier; 2) the interference I_{lj-} from the nodes without storing content j in the l -th tier. The positions of the nodes in I_{lj+} can be modeled as an HPPP Φ_{lj+} with density $p_{lj} \lambda_l$ while the nodes in I_{lj-} are also located as an independent HPPP Φ_{lj-} with density $(1-p_{lj}) \lambda_l$. Considering the condition $d_{l,k} \geq r_{l|j} \geq r \left(\frac{S_l}{S_i} \right)^{\frac{1}{\beta}}$, we thus have

$$\begin{aligned}
\mathcal{L}_{I_{lj+}}(\tau S_i^{-1} r^\beta) &= \mathbb{E}_{I_{lj+}} \left[e^{-\tau S_i^{-1} r^\beta I_{lj+}} \right] = \mathbb{E}_{\Phi_{lj+}, |h_{l,k}|^2} \left[e^{-\tau S_i^{-1} r^\beta \sum_{k \in \Phi_{lj+} \setminus \{n_{io}\}} S_l |h_{l,k}|^2 d_{l,k}^{-\beta}} \right] \\
&= \mathbb{E}_{\Phi_{lj+}} \left[\prod_{k \in \Phi_{lj+} \setminus \{n_{io}\}} \frac{1}{1 + \tau S_i^{-1} r^\beta S_l d_{l,k}^{-\beta}} \right] \\
&\stackrel{(b)}{=} \exp \left(-2\pi p_{lj} \lambda_l \int_{r \left(\frac{S_l}{S_i} \right)^{\frac{1}{\beta}}}^\infty \left(1 - \frac{1}{1 + \tau S_i^{-1} r^\beta S_l x^{-\beta}} \right) x dx \right) \\
&\stackrel{(c)}{=} \exp \left(-\pi p_{lj} \lambda_l \tau^{\frac{2}{\beta}} \left(\frac{S_l}{S_i} \right)^{\frac{2}{\beta}} r^2 \frac{2}{\beta} \int_{\tau^{-1}}^\infty \left(\frac{\mu^{\frac{2}{\beta}-1}}{1+\mu} \right) d\mu \right) \\
&= \exp \left(-\pi p_{lj} \lambda_l \left(\frac{S_l}{S_i} \right)^{\frac{2}{\beta}} r^2 H(\tau, \beta) \right), \tag{32}
\end{aligned}$$

where (b) follows from the probability generating functional (PGFL) of the PPP and (c) is obtained using $\mu = \tau^{-1} S_i r^{-\beta} S_l^{-1} x^\beta$, $H(\tau, \beta)$ denotes $\frac{2\tau}{\beta-2} {}_2F_1(1, 1 - \frac{2}{\beta}, 2 - \frac{2}{\beta}, -\tau)$. Similarly,

let $D(\tau, \beta)$ be $\frac{2}{\beta}\tau^{\frac{2}{\beta}}B(\frac{2}{\beta}, 1 - \frac{2}{\beta})$, the Laplace transform $\mathcal{L}_{I_{lj-}}$ is

$$\begin{aligned}\mathcal{L}_{I_{lj-}}(\tau S_i^{-1}r^\beta) &= \mathbb{E}_{I_{lj-}}[\exp(-\tau S_i^{-1}r^\beta I_{lj-})] \\ &= \exp\left(-(1-p_{lj})\lambda_l \int_{R^2} \left(1 - \frac{1}{1 + \tau S_i^{-1}r^\beta S_l x^{-\beta}}\right) x dx\right) \\ &= \exp\left(-\pi(1-p_{lj})\lambda_l \left(\frac{S_l}{S_i}\right)^{\frac{2}{\beta}} r^2 D(\tau, \beta)\right).\end{aligned}\quad (33)$$

Substituting (7), (32), (33) into (31), we have Proposition 1.

B. Proof of Proposition 2

It is easy to see that the feasible set $\{\mathbf{P} | 0 \leq p_{ij} \leq 1, \sum_{j=1}^M p_{ij} \leq Q_i, \forall i \in \mathcal{N}, \forall j \in \mathcal{M}\}$ is a convex set. The objective function $\mathcal{C}'(\mathbf{P})$ is a summation of M polynomials. For $\forall j \in \mathcal{M}$, the j -th polynomial is given by $F_j = \frac{\sum_{i=1}^N V_{ij} p_{ij}}{\sum_{i=1}^N G_i p_{ij} + E}$, where $V_{ij} = \lambda_i S_i^{\frac{2}{\beta}} t_j$, $G_i = \lambda_i S_i^{\frac{2}{\beta}} T(\tau, \beta)$, and $E = D(\tau, \beta) \sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}}$. The first derivative of F_j is

$$\frac{\partial F_j}{\partial p_{ij}} = \frac{\partial \mathcal{C}'}{\partial p_{ij}} = \frac{V_{ij} E}{\left(\sum_{i=1}^N G_i p_{ij} + E\right)^2} > 0. \quad (34)$$

We then have

$$\frac{\partial^2 F_j}{\partial p_{ij} \partial p_{kj}} = -m a_i a_k < 0, \quad (35)$$

where $m = \frac{2T(\tau, \beta)t_j E}{(\sum_{i=1}^N G_i p_{ij} + E)^3}$, $a_i = \lambda_i S_i^{\frac{2}{\beta}}$, and $a_k = \lambda_k S_k^{\frac{2}{\beta}}$. Then, the Hessian matrix $\mathbf{H}_{N \times N}$ of F_j is given by

$$\mathbf{H}_{N \times N} \triangleq -m[a_i a_k]_{N \times N} = -m(\mathbf{a}\mathbf{a}^T), \quad (36)$$

where we define the column vector $\mathbf{a} \triangleq (a_1, a_2, \dots, a_N)^T$. Obviously, the Hessian matrix \mathbf{H} is a real symmetric negative semi-definite matrix, i.e., F_j is concave of p_{ij} . According to the property that the summation of concave functions is also a concave function, \mathcal{C}' is also a concave function of \mathbf{P} . Thus, Problem P1 is a concave optimization problem.

C. Proof of Lemma 3

By constructing the Lagrangian function of P2, we have

$$L(\mathbf{P}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\eta}) = \mathcal{C}'(\mathbf{P}) + \sum_{i=1}^N \sum_{j=1}^M \mu_{ij} p_{ij} + \sum_{i=1}^N \sum_{j=1}^M \lambda_{ij} (1 - p_{ij}) + \sum_{i=1}^N \eta_i (Q_i - \sum_{j=1}^M p_{ij}), \quad (37)$$

where μ_{ij} , λ_{ij} are the Lagrange multipliers associated with (1), η_i is the Lagrange multiplier associated with (12). And we use $\boldsymbol{\mu} \triangleq (\mu_{ij})_{N \times M}$, $\boldsymbol{\lambda} \triangleq (\lambda_{ij})_{N \times M}$, $\boldsymbol{\eta} \triangleq (\eta_i)_{i \in \mathcal{N}}$. Thus, we have

$$\frac{\partial L(\mathbf{P}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\eta})}{\partial p_{ij}} = \frac{V_{ij}E}{(\sum_{k=1}^N G_k p_{kj} + E)^2} + \mu_{ij} - \lambda_{ij} - \eta_i. \quad (38)$$

The KKT conditions are then written as follows

$$\frac{\partial L(\mathbf{P}^*, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\eta})}{\partial p_{ij}^*} = 0, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M}, \quad (39)$$

$$\mu_{ij} p_{ij}^* = 0, \quad \lambda_{ij}(1 - p_{ij}^*) = 0, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M}, \quad (40)$$

$$\eta_i(Q_i - \sum_{j=1}^M p_{ij}^*) = 0, \quad \forall i \in \mathcal{N}, \quad (41)$$

$$\sum_{j=1}^M p_{ij}^* = Q_i, \quad 0 \leq p_{ij}^* \leq 1, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \quad (42)$$

$$\mu_{ij} \geq 0, \quad \lambda_{ij} \geq 0, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M}. \quad (43)$$

According to (38) and (39), we have that

$$\eta_i = \frac{V_{ij}E}{(\sum_{k=1}^N G_k p_{kj}^* + E)^2} + \mu_{ij} - \lambda_{ij} \quad \forall k \in \mathcal{N}, \forall j \in \mathcal{M}. \quad (44)$$

As a result, we consider the following three cases to analyze the optimal solution:

- $\eta_i \geq \frac{V_{ij}E}{(\sum_{k=1, \neq i}^N G_k p_{kj}^* + E)^2}$: If $0 < p_{ij}^* \leq 1$, then we have $0 \leq \lambda_{ij} < \mu_{ij}$, which is in conflict with $\mu_{ij} = 0$ due to (40). If $p_{ij}^* = 0$, we have $\lambda_{ij} = 0$ and $\mu_{ij} \geq 0$. Thus, we have $p_{ij}^* = 0$.
- $\frac{V_{ij}E}{(\sum_{k=1, \neq i}^N G_k p_{kj}^* + G_i + E)^2} < \eta_i < \frac{V_{ij}E}{(\sum_{k=1, \neq i}^N G_k p_{kj}^* + E)^2}$: If $0 < p_{ij}^* < 1$, we have $\mu_{ij} = \lambda_{ij} = 0$ due to (40); If $p_{ij}^* = 0$, we have $0 \leq \mu_{ij} < \lambda_{ij}$, which is in conflict with $\lambda_{ij} = 0$ due to (40); If $p_{ij}^* = 1$, we have $0 \leq \lambda_{ij} < \mu_{ij}$ and $\mu_{ij} = 0$, which are contradictory. Thus, we have $0 < p_{ij}^* < 1$ and $p_{ij}^* = \frac{1}{G_i} \left(\sum_{k=1}^N G_k p_{kj}^* = \sqrt{\frac{V_{ij}E}{\eta_i}} - E \right)$.
- $\eta_i \leq \frac{V_{ij}E}{(\sum_{k=1, \neq i}^N G_k p_{kj}^* + G_i + E)^2}$: Similar to the analysis for the first case, we have $p_{ij}^* = 1$ and $\mu_{ij} = 0, \lambda_{ij} \geq 0$.

Summarizing above results, we thus have (13). Based on (42), we have that η_i satisfies $\sum_{j=1}^M p_{ij}^* = Q_i$ for $\forall i \in \mathcal{N}$. Thus, we complete the proof.

D. Proof of Corollary 3

- For $\forall i, j \in \mathcal{M}$ and $i < j$, we have $t_i \geq t_j$. Then according to (14), there are only three possible cases: $p_i^* = p_j^* = 1$, $p_i^* > p_j^*$, $p_i^* = p_j^* = 0$, which can be summarized as $p_i^* \geq p_j^*$.

Therefore, p_j^* decreases with the index j , i.e., p_j^* increases with the t_j .

- Since $1 \leq Q \ll M$, we assume $p_i^* = 1$ for $1 \leq i \leq n$ and $p_j^* = 0$ for $m \leq j \leq M$ ($1 \leq n < m < M$). Thus, we have $\sum_{j=n+1}^{m-1} \left(\frac{1}{T(\tau, \beta)} \sqrt{\frac{t_j D(\tau, \beta)}{\eta^*}} - \frac{D(\tau, \beta)}{T(\tau, \beta)} \right) = Q - n$. When the cache size Q increases to Q' , η^* will decrease to a new value $\eta^{*'}$ to satisfy $\sum_{j=1}^M p_j^{*'} = Q'$ and $\frac{1}{T(\tau, \beta)} \sqrt{\frac{t_j D(\tau, \beta)}{\eta^*}} - \frac{D(\tau, \beta)}{T(\tau, \beta)}$ will increase to $\frac{1}{T(\tau, \beta)} \sqrt{\frac{t_j D(\tau, \beta)}{\eta^{*'}}} - \frac{D(\tau, \beta)}{T(\tau, \beta)}$. Thus, if $1 \leq i \leq n$, we have $p_j^{*'} = p_j^* = 1$; if $n+1 \leq i \leq m-1$, we have $p_j^{*'} > p_j^*$; if $m \leq j \leq M$, we have $p_j^{*'} = p_j^* = 0$ or $p_j^{*'} > p_j^* = 0$. Summarizing above three cases, we have $p_j^* \leq p_j^{*'}$, i.e., p_j^* increase with cache size Q .

E. Proof of Proposition 3

By letting $x_j = \frac{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}} p_{ij}}{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}}}$, it is clear that the objective function of P2 is the same as that of P3. Hence, to prove the first part of proposition it suffices to show that the feasible set of P2 is a subset of that of P3. Denote $\mathcal{P} = \{\mathbf{P} = [p_{ij}]_{N \times M} \mid 0 \leq p_{ij} \leq 1, \sum_{j=1}^M p_{ij} = Q_i, \forall i \in \mathcal{N}, \forall j \in \mathcal{M}\}$ as the feasible set of P2 and $\mathcal{X} = \{\mathbf{x} = (x_j)_{1 \times M} \mid 0 \leq x_j \leq 1, \sum_{j=1}^M x_j = \frac{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}} Q_i}{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}}}, \forall j \in \mathcal{M}\}$ as the feasible set of P3. Let $\mathbf{a} = (\frac{\lambda_i S_i^{\frac{2}{\beta}}}{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}}})_{1 \times N}$, then for any given element $\mathbf{P} \in \mathcal{P}$, we can define $\mathbf{x} = \mathbf{aP}$. Since $\frac{\partial x_j}{\partial p_{ij}} = \frac{\lambda_i S_i^{\frac{2}{\beta}}}{\sum_{l=1}^N \lambda_l S_l^{\frac{2}{\beta}}} > 0$ and $0 \leq p_{ij} \leq 1$ for $\forall i \in \mathcal{N}, \forall j \in \mathcal{M}$, we have $0 \leq x_j \leq 1$ for $\forall j \in \mathcal{M}$. Due to $\sum_{j=1}^M p_{ij} = Q_i$ for $i \in \mathcal{N}$, we have $\sum_{j=1}^M x_j = \frac{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}} \sum_{j=1}^M p_{ij}}{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}}} = \frac{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}} Q_i}{\sum_{i=1}^N \lambda_i S_i^{\frac{2}{\beta}}}$. Thus, we have $\mathbf{x} \in \mathcal{X}$. This means that for any $\mathbf{P} \in \mathcal{P}$, it always holds that $\mathbf{x} = \mathbf{aP} \in \mathcal{X}$. Hence, the feasible set \mathcal{X} of P3 includes the transformation of the feasible set \mathcal{P} of P2 as a subset. As a result, the optimal value of P3 provides an upper bound of that of P2.

We next prove the second part of the proposition. If $Q_i = Q$ for $\forall i \in \mathcal{N}$, then for any given element $\mathbf{x} \in \mathcal{X}$, we can always construct an \mathbf{P}' where $p_{ij} = x_j$ for $i \in \mathcal{N}$. Due to $0 \leq p_{ij} = x_j \leq 1$ for $\forall i \in \mathcal{N}, \forall j \in \mathcal{M}$ and $\sum_{j=1}^M p_{ij} = \sum_{j=1}^M x_j = Q$ for $\forall i \in \mathcal{N}$, we thus have $\mathbf{P}' \in \mathcal{P}$. This suggests that if $\mathbf{x}^* \in \mathcal{X}$ is an optimal solution of P3, then \mathbf{P}' with $p_{ij} = x_j^*$ is also an optimal solution of P2. Hence, the optimal values of P2 and P3 are equal when $Q_i = Q, \forall i \in \mathcal{N}$. The proposition is thus proved.

F. Proof of Theorem 1

According to (19), when $\mathcal{C}'_1(\mathbf{p}_{1,e})$ is maintained constant and λ_j , S_j and Q_j are fixed values for $\forall j \in \mathcal{N}, j \neq i$, the equivalent cache size Q_e also remains unchanged. We only need to consider the interactions of the i -th tier's parameters, thus we can derive that

$$\lambda_i S_i^{\frac{2}{\beta}} = \frac{Q_e \sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}} - \sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}} Q_j}{Q_i - Q_e}. \quad (45)$$

Substituting (22) into (45), (23) into (45) respectively, we have (20) , (21).

G. Proof of Corollary 6

Equation (45) can be rewritten as following

$$\lambda_i S_i^{\frac{2}{\beta}} = \frac{\left(Q_e - \frac{\sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}} Q_j}{\sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}}} \right) \sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}}}{Q_i - Q_e}. \quad (46)$$

Obviously, when $Q_e > \frac{\sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}} Q_j}{\sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}}}$, we have $K_1 > 0$. Based on $Q_e = \frac{\lambda_i(S_i)^{\frac{2}{\beta}} Q_i + \sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}} Q_j}{\lambda_i(S_i)^{\frac{2}{\beta}} + \sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}}}$, $Q_e > \frac{\sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}} Q_j}{\sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}}}$ is equivalent to $Q_i > \frac{\sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}} Q_j}{\sum_{j=1, \neq i}^N \lambda_j S_j^{\frac{2}{\beta}}}$. Meanwhile, because of $\lambda_i(S_i)^{\frac{1}{\beta}} > 0$, we have $Q_i > Q_e$. As a result, we have $K_1 > 0$ and $Q_i - Q_e > 0$ when $Q_i \in (Q_e, +\infty)$, then $K_1 < 0$ and $Q_i - Q_e < 0$ when $Q_i \in [0, Q_e)$. Because the relationship of λ_i and Q_i is inversely proportional, thus λ_i decreases as Q_i increases when $K_1 > 0$ and $Q_i - Q_e > 0$, while λ_i increases as Q_i increases when $K_1 < 0$ and $Q_i - Q_e < 0$. Similarly, we have S_i decreases as Q_i increases when $K_2 > 0$ and $Q_i - Q_e > 0$, while S_i increases as Q_i increases when $K_2 < 0$ and $Q_i - Q_e < 0$. Finally, we have the corollary.

REFERENCES

- [1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update 2015-2020," *white paper*, Feb. 2016.
- [2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5g be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [3] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic Geometry for Modeling, Analysis, and Design of Multi-Tier and Cognitive Cellular Wireless Networks: A Survey," *IEEE Commun. Surveys and Tutorials*, vol. 15, no. 3, pp. 996–1019, Jun. 2013.
- [4] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in The Air: Exploiting Content Caching and Delivery Techniques for 5G Systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.
- [5] E. Baştuğ, M. Bennis, and M. Debbah, "Living on the Edge: The Role of Proactive Caching in 5G Wireless Networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug. 2014.

- [6] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, 2014.
- [7] H. Liu, Z. Chen, and L. Qian, "The three primary colors of mobile systems," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 15–21, September 2016.
- [8] H. Liu, Z. Chen, X. Tian, X. Wang, and M. Tao, "On content-centric wireless delivery networks," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 118–125, Dec. 2014.
- [9] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, 2011.
- [10] E. Baştuğ, M. Bennis, M. Kountouris, and M. Debbah, "Cache-enabled small cell networks: Modeling and tradeoffs," *EURASIP Journal on Wireless Commun. and Networking*, vol. 2015, no. 1, p. 4, 2015.
- [11] S. Tamoor-ul Hassan, M. Bennis, P. H. J. Nardelli, and M. Latva-Aho, "Caching in wireless small cell networks: A storage-bandwidth tradeoff," *IEEE Commun. Letters*, vol. 20, no. 6, pp. 1175–1178, 2016.
- [12] C. Yang, Y. Yao, Z. Chen, and B. Xia, "Analysis on cache-enabled wireless heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 131–145, 2016.
- [13] S. Tamoor-ul Hassan, M. Bennis, P. H. J. Nardelli, and M. Latva-Aho, "Modeling and analysis of content caching in wireless small cell networks," *IEEE ISWCS*, 2015.
- [14] Z. Chen, J. Lee, T. Q. S. Quek, and M. Kountouris, "Cooperative caching and transmission design in cluster-centric small cell networks," <http://arxiv.org/abs/1601.00321>, 2016.
- [15] K. Shanmugam, N. Golrezaei, A. Dimakis, A. Molisch, and G. Caire, "FemtoCaching: Wireless Content Delivery Through Distributed Caching Helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013.
- [16] K. Poularakis, G. Iosifidis, and L. Tassiulas, "Approximation algorithms for mobile data caching in small cell networks," *IEEE Trans. Commun.*, vol. 62, no. 10, pp. 3665–3677, Oct. 2014.
- [17] B. Blaszczyszyn and A. Giovanidis, "Optimal geographic caching in cellular networks," in *proc. IEEE ICC*, 2015.
- [18] Y. Chen, M. Ding, J. Li, Z. Lin, G. Mao, and L. Hanzo, "Probabilistic small-cell caching: Performance analysis and optimization," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2016.
- [19] S. H. Chae and W. Choi, "Caching placement in stochastic wireless caching helper networks: Channel selection diversity via caching," *IEEE Trans. Wireless Commun.*, vol. PP, no. 99, pp. 1–1, 2016.
- [20] J. Song, H. Song, and W. Choi, "Optimal caching placement for wireless femto-caching network," [Online]. Available: [arXiv preprint arXiv:1605.09519](https://arxiv.org/abs/1605.09519), 2016.
- [21] D. Liu and C. Yang, "Caching policy toward maximal success probability and area spectral efficiency of cache-enabled hetnets," [Online]. Available: [arXiv preprint arXiv:1608.03749](https://arxiv.org/abs/1608.03749), 2016.
- [22] B. Serbetci and J. Goseling, "On optimal geographical caching in heterogeneous cellular networks," [Online]. Available: [arXiv preprint arXiv:1601.07322](https://arxiv.org/abs/1601.07322), 2016.
- [23] A. Liu and V. K. N. Lau, "How much cache is needed to achieve linear capacity scaling in backhaul-limited dense wireless networks?" *IEEE/ACM Transactions on Networking*, vol. PP, no. 99, pp. 1–10, 2016.
- [24] H.-S. Jo, Y. J. Sang, P. Xia, and J. Andrews, "Heterogeneous Cellular Networks with Flexible Cell Association: A Comprehensive Downlink SINR Analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, Oct. 2012.
- [25] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.